Экзамен

Нужно ответить на 6 вопросов под номерами, которые вы получили в результате запуска датчика случайных чисел. Ответ на каждый вопрос оценивается, исходя из максимума в 2 балла. Если Вы набрали более 10 баллов, то ставится оценка 10.

1. Дайте определение выборочной медианы. Сформулируйте хотя бы одно утверждение, показывающее связь между выборочным средним и выборочной медианой.

2. Объясните метод построения статистической оценки с меньшим смещением при помощи Jackknife.

3. Объясните метод построения рандомизированных равномерно наиболее мощных тестов. В каких ситуациях следует отказаться от обычных (нерандомизированных) тестов и использовать рандомизированные?

4. Сформулируйте утверждение, показывающее связь между эмпирическим коэффициентов корреляции Пирсона и коэффициентов детерминации R^2.

5. Дайте определение гребневой регрессии (регрессии ридж). В какой ситуации следует использовать модель гребневой регрессии, а не обычную линейную модель ?

6. Сформулируйте теорему Гаусса-Маркова. Приведите интерпретацию этой теоремы в контексте разложения среднеквадратической ошибки на смещение и дисперсию (bias-variance decomposition).

7. Дайте определение регрессограммы. Дайте определение линейного сглаживателя (linear smoother) и объясните, почему регрессограмма является линейным сглаживателем. Дайте определение эффективного количества степеней свободы (effective degrees of freedom) и найдите значение этой характеристики для регрессограммы.

8. Дайте определение функции выживаемости (survival function). Опишите метод оценивания этой функции через построение таблицы дожития (life-table estimates) и метод Каплана-Мейера.

9. Дайте определение функции риска (hazard function) в анализе выживаемости. Опишите метод оценивания этой функции через построение таблицы дожития (life-table estimates) и метод Каплана-Мейера.

10. Опишите логранговый тест (log-rank test) для сравнения двух групп (двух кривых выживаемости).

11. Дайте определение экспоненциального семейства распределений. Приведите 2 примера экспоненциальных семейств распределений.

12. Докажите, что применение метода моментов и метода максимального правдоподобия для экспоненциального семейства распределений приводит к одинаковым оценкам.

13. Объясните, как строится график квантиль-квантиль. Объясните, почему в случае нормального распределения выборки точки на этом графике лежат на одной прямой.

14. Объясните, что такое bias-variance decomposition. Покажите суть этого явления на примере гистограммы.

15. В каком смысле гистограмма является оптимальной оценкой плотности?

16. В каком смысле ядерная оценка плотности (kernel density estimate) является оптимальной оценкой плотности?

17. В чём состоит основная идея использования методов "nrd" и "nrd0" для выбора параметра bandwidth ядерной оценки плотности?

18. Сформулируйте теоретический результат, показывающий, что ядерные оценки плотности, построенные при помощи ядер с переменным знаком (то есть, при помощи ядер, которые принимают как положительные, так и отрицательные значения), обладают лучшими асимптотическими свойствами, чем ядерные оценки плотности, построенные при помощи неотрицательных ядер.

19. Объясните суть метода кросс-проверки для выбора параметров оценок плотности распределения.

20. Объясните суть EM-алгоритма.

21. Сформулируйте лемму Неймана-Пирсона.

22. Сформулируйте теорему Уилкса.

23. Сформулируйте теорему Пирсона и объясните, как эта теорема может быть использована для тестирования гипотезы о том, что выборка имеет заданное распределение.

24. Сформулируйте теоретический результат, из которого следует метод для проверки гипотезы о равенстве нулю теоретического коэффициента корреляции Пирсона.

25. Объясните, почему для независимых случайных величин, имеющих непрерывные распределения, теоретический коэффициент корреляции Кендалла равен 0.

26. Сформулируйте теорему об асимптотическом поведении коэффициента корреляции Спирмена и объясните, каким образом можно использовать эту теорему для тестирования гипотезы о независимости непрерывных переменных.

27. Для чего нужно разложение Эджворта (Edgworth expansion)? Сформулируйте соответствующий результат о сходимости суммы независимых одинаково распределённых случайных величин с нулевым средним и единичной дисперсией.

28. Опишите алгоритм метода LOESS.

29. Объясните, метод построения оценки Надарая-Ватсона.

30. Объясните суть критерия Манна-Уитни для независимых групп.

31. Опишите схему применения теста Уилкоксона для парных повторных наблюдений.

32. Опишите модель, в которой используется тест Фридмана для k зависимых выборок, и сформулируйте теоретический результат, который лежит в основе этого теста.

33. Объясните математическую идею, лежащую в основе дисперсионного анализа (метод ANOVA).

34. Сформулируйте теорему, которая лежит в основе статистических тестов в модели линейной регрессии (тестов для проверки значимости коэффициентов и значимости статистики R-квадрат).

35. Сформулируйте теоретический результат, который лежит в основе метода обобщённой кросс-валидации для моделей регрессии. Объясните суть метода.

36. Опишите схему построения обобщённых линейных моделей (GLM). Объясните, почему логистическая регрессия является частным случаем обобщённой линейной модели.

37. Объясните, что такое null deviance и residual deviance в результатах работы функции glm в языке R. Каким образом можно использовать данные величины для анализа качества построенной модели?

38. Объясните метод определения качества модели классификации при помощи ROC AUC.

39. Дайте определение базиса Хаара. В каком пространстве этот набор функций является базисом?

40. Дайте определение величин weights of evidence и information value. Объясните, почему величина information value является неотрицательной.

41. В чём принципиальное отличие между методами параметрической и непараметрической статистики?